Determination of changes in streamflow variance by means of a wavelet-based test

Anthony T. Cahill

Department of Civil Engineering, Texas A&M University, College Station, Texas, USA

Received 8 January 2001; revised 19 December 2001; accepted 19 December 2001; published 1 June 2002.

[1] Changes in the variance of streamflow over time may possibly be caused by land cover changes or climate changes. Knowledge of these possible changes in the time series structure is desirable for validation of modeling efforts; detection is not always straightforward, however. We present the use of a wavelet-based statistical test for the detection of changes in the variance of a time series applied to records of streamflow from 12 rivers in the United States from the period 1954–1999. The test is able to show during which time periods and at what wavelet scales changes in the variance are taking place. It is found that the major similarities in variance change for all the streamflow time series exist when the largest time domain wavelet is used and also when a wavelet of the second largest frequency scale is used. Sampling considerations prevent exact localization of the time and frequency of the change(s). Results of the test indicate that for the data analyzed in this study a change in the variance of the streamflow time series most likely did take place over the time period examined. *INDEX TERMS:* 1833 Hydrology: Hydroclimatology; 1860 Hydrology: Runoff and streamflow; 1869 Hydrology: Stochastic processes; *KEYWORDS:* wavelets, streamflow, climate change

1. Introduction

[2] Estimation and description of daily streamflow remain one of the problems of interest in surface water hydrology. New approaches to this problem continue to be developed [e.g., *Liu et al.*, 1998]. This interest is driven by not only the widespread use of riparian water for human water supply but also the integrating effect of the hydrologic state of a watershed evinced by streamflow. The amount and timing of streamflow are related to the amount of rainfall, soil moisture, and land cover of the watershed and hence may yield information on changes in land cover or climate. Recent analysis of twentieth century time series of precipitation and evaporation have indicated that the hydrologic cycle may be becoming more intense, with more precipitation, evaporation, and runoff [*Karl et al.*, 1996; *Brutsaert and Parlange*, 1998; *Kiely et al.*, 1998]. The forward problem, changes in land cover and/or climate regime leading to changes in the magnitude and timing of river flow, has been explored through basin-scale simulations, with an aim of evaluating the potential effects of global climate change on a regional hydrologic cycle [*Hamlet and Lettenmaier*, 1999; *Matheussen et al.*, 2000].

[3] In order to validate these and future forward models that estimate the hydrologic effect of land cover and climatic change it may prove useful to investigate the statistical characteristics of actual streamflow time series and evaluate their changes in time. Although we would like forward models to reproduce accurately the entire probability distribution of streamflows under a set of land surface and climatic conditions, in practice, validation involves matching the first two statistical moments, and so we restrict our discussion to them. The possibility of changes in mean streamflow has been evaluated using traditional trend and change point analysis by *Changnon and Demissie* [1996]. The treatment of changes in the variance structure of a time series is somewhat more difficult and has been the subject of a number of approaches in the statistical literature. Proposed methods include cumulative sum of squares [*Inclán and Tiao*, 1994], Bayesian inference [*McCulloch and Tsay*, 1993], periodogram tests [*Picard*, 1985], local cosine bases [*Mallat et al.*, 1998], and time-correlation analysis [*Li*, 1998], among others. In this study, we propose to examine a number of daily streamflow records for evidence of change in the variance using a wavelet-based statistical test of the periodogram.

[4] It should be noted that a change in the variance of a streamflow time series is not necessarily problematic for streamflow modeling efforts. The traditional Box and Jenkins autoregressive moving average (ARMA) approach, often used for streamflow modeling, requires a stationary Gaussian time series [*Bras and Rodriguez-Iturbe*, 1993]. Since time series of streamflow are, in general, nonstationary in the mean and also non-Gaussian (streamflow is strictly positive, and extreme flows can lead to heavy tailed distributions), generally, the streamflow time series is first transformed using the Box-Cox approach [*Box and Cox*, 1964]:

$$z_t = \begin{cases} \frac{y_t^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \log\, y_t & \lambda = 0, \end{cases} \tag{1}$$

where λ is the transform parameter which turns the non-Gaussian time series y_t into a (nearly) Gaussian one, z_t. Next, differencing is used to remove nonstationarity in the mean. Long-memory effects in hydrologic time series are addressed by use of fractional differencing. Maximum likelihood methods exist to estimate all of the needed parameters [*Johnson and Wichern*, 1998; *Beran*, 1995]. What is lost in these standard transformations is any information about when the nonstationarity in the parameters occurs. For example, since the maximum likelihood value of λ is determined by minimizing the sum over all the residuals, without regard to their location in the time series, the resulting transformation parameter can tell us nothing about if or when a change in the variance structure occurred. Since our interest in this study is whether a change in the variance in streamflow occurs and when that change occurs, we will focus our analysis on the untransformed time series of streamflow. Once the location of a change

0043-1397/02/2000WR000192$09.00

Table 1. List of USGS Gauging Stations Used in This Paper and Their Drainage Areas

Site	Location	USGS Station Number	Drainage Area, km^2
1	Connecticut River at Thompsonville, Connecticut	01184000	25,010
2	Schoharie Creek at Prattsville, New York	01350000	614
3	Esopus Creek at Coldbrook, New York	01362500	497
4	Rapidan River near Culpeper, Virginia	01667500	1,222
5	Sabine River near Ruliff, Texas	08030500	24,153
6	West Fork Trinity River at Grand Prairie, Texas	08049500	7,935
7	Colorado River at Columbus, Texas	08161000	107,806
8	San Marcos River at Luling, Texas	08172000	2,170
9	Pecos River near Girvin, Texas	08446500	76,531
10	Klamath River at Keno, Oregon	11509500	10,149
11	Nisqually River at La Grande, Washington	12086500	756
12	Willamette River at Albany, Oregon	14174000	12,531

point is known, standard modeling techniques can be used for each separate section of the time series.

[5] Wavelets have previously been used to analyze streamflow time series by using plots of the wavelet coefficients as functions of time and frequency (wavelet scalogram) [*Smith et al.*, 1998]. Scalograms of streamflow calculated using the Mexican hat wavelet were compared for 91 different rivers. It was found that the scalograms of the streamflow time series exhibited characteristics which would allow the rivers to be grouped into five different categories, depending on what sort of input (snowmelt, brief intense storms, etc.) drove streamflow. Wavelet analysis has also been applied to a number of different environmental phenomena, including rainfall fields [*Kumar and Foufoula-Georgiou*, 1993], climate ''wetness'' index [*Jiang et al.*, 1997], the Southern Oscillation signal [*Wang and Wang*, 1996], sea level fluctuations [*Percival and Mofjeld*, 1997], the relationship between the Madden-Julian oscillation and the El Niño–Southern Oscillation (ENSO) signal [*Whitcher et al.*, 2000], and boundary layer turbulence [*Katul and Parlange*, 1995].

[6] Because of their ability to decompose a signal with respect to both time and frequency, there have been a number of different investigations into the use of wavelets for the determination of the second-order stationarity of a (possibly nonstationary) time series. One possible method considers the wavelet variance itself and examines its evolution through time. The wavelet variance used in this approach is that defined by the maximum overlap discrete wavelet transform (MODWT), which has also been called the undecimated discrete wavelet transform (DWT) or the translation-invariant DWT [*Coifman and Donoho*, 1995]. The MODWT basically performs a DWT without the reduction of the wavelet coefficients by a power of 2 at each step. As pointed out by *Percival* [1995], the MODWT is superior to the standard discrete wavelet transform with respect to estimation of the wavelet variance. The wavelet variance at each decomposition level can be computed as the sum of the wavelet coefficients at that level. In order to test for stationarity of this variance it is useful to examine the cumulative variance, for which nonparametric tests can be developed [*Percival and Walden*, 2000]. This approach has been used to examine the time series of annual low flows in the Nile River for changes in the variance [*Whitcher et al.*, 1998]. Another approach to using wavelets to examine the variance of possibly nonstationary time series extends the techniques of singular-spectrum analysis by using varying window widths at different correlation orders to obtain data-adaptive wavelet transforms that highlight nonstationary dominant structures [*Yiou et al.*, 2000].

[7] An alternate approach to the detection of nonstationarity of the variance using wavelets involves examination of the local periodogram, which is simply the periodogram calculated for some subsection of the entire time series [*Neumann and von Sachs*, 1997; *von Sachs and Neumann*, 2000]. Given the equivalence of

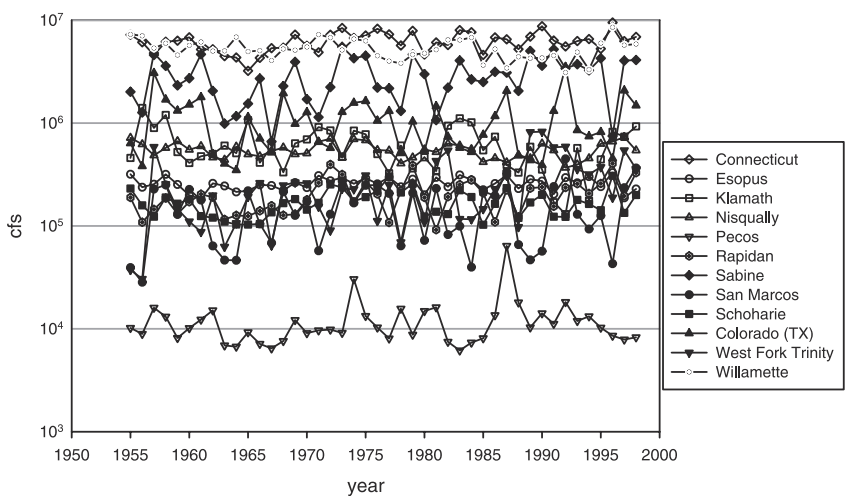


Figure 1. Total annual flow for the 12 rivers used in this study. A logarithmic scale is used only so that the rivers with low annual flow can be resolved; the trend fitting discussed in the text was done against untransformed data.

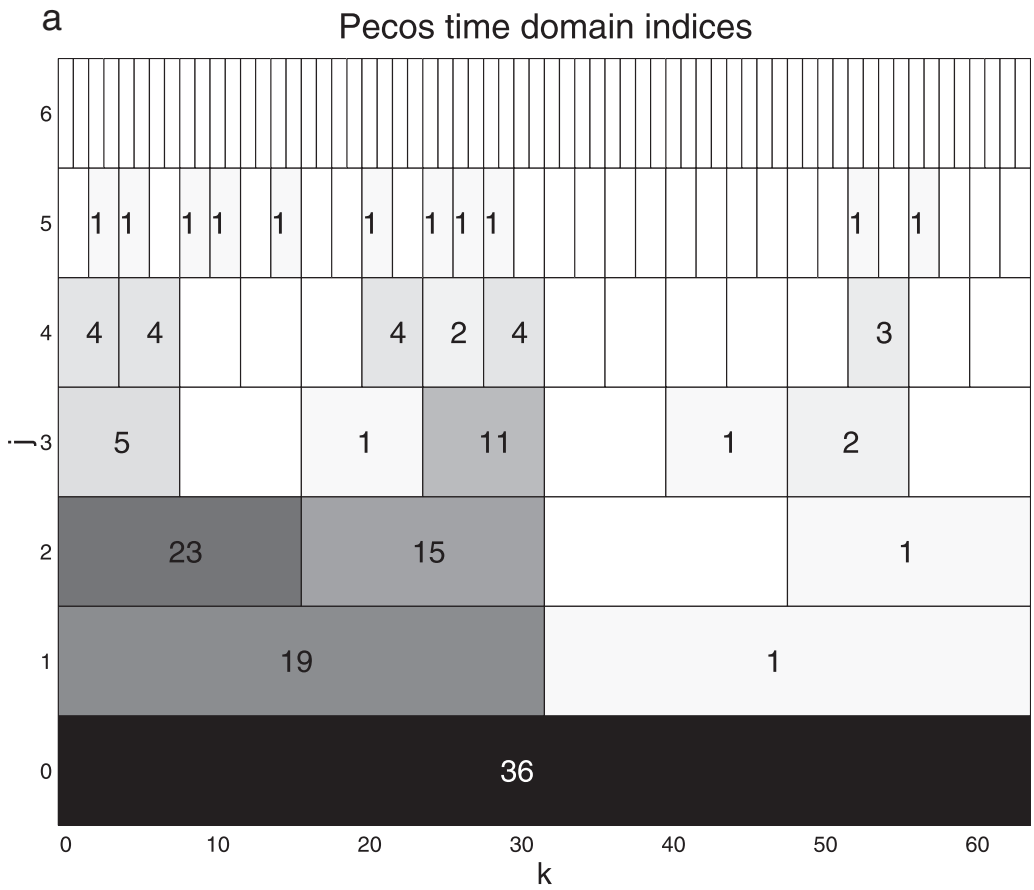


Figure 2. Plots of the number of statistically significant values of $\tilde{\alpha}_{j,k;j',k'}$ for all values of j', k' (frequency indices) at a given pair of j, k (time indices) for the (a) Pecos, (b) Rapidan, and (c) Willamette Rivers. Statistical significance indicates that the periodogram (and hence the variance) is nonconstant for given set of (j, k, j', k'). The number of statistically significant values has been placed in the nonzero cells since the gray scale is close to white when this number is close to 0. The same gray scale is used for Figures 2a–2c.

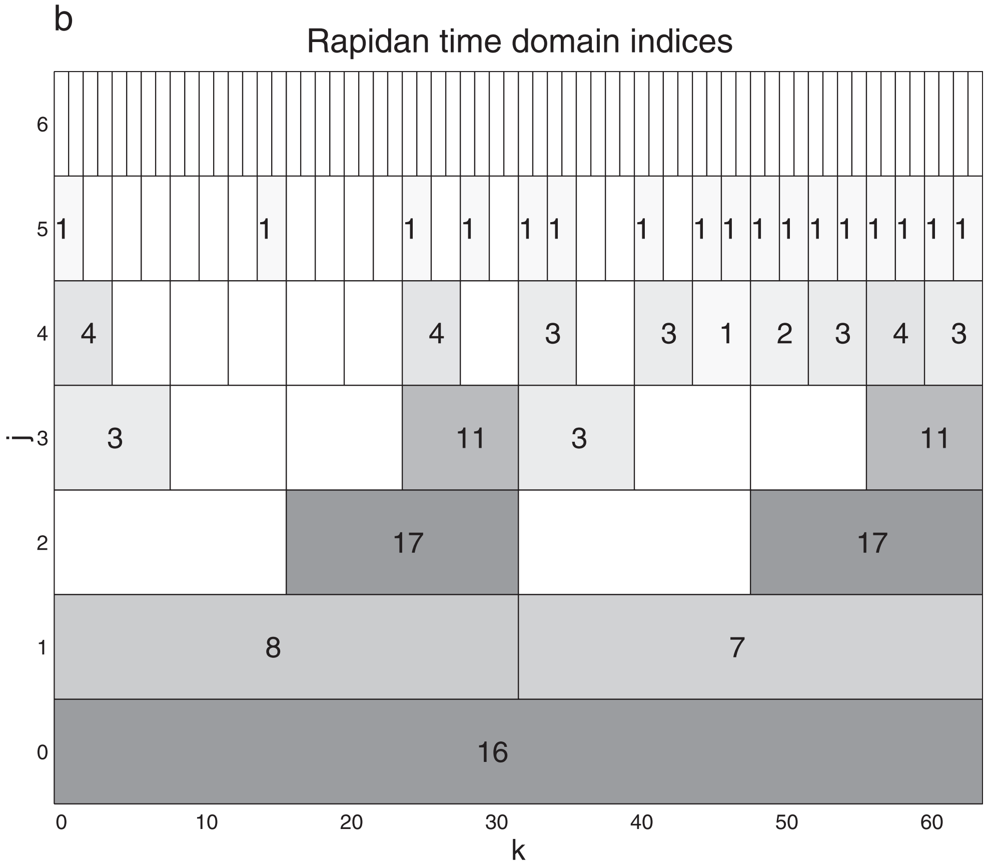
the variance and spectrum with respect to Fourier transformation, it can be seen heuristically that a change in the local variance will be the same as a change in the local spectrum. A time-varying spectral density of a time series can be defined as $f(u, \omega)$, where u is a time index and ω is a frequency index. For the null hypothesis of a stationary time series we have $H_0 : f(u, \omega) = f(\omega)$; the spectrum depends only on the wave number. In this approach, wavelets are used as test functions to develop a set of coefficients which depend on the scale and translation of the wavelet and which, under the null hypothesis, ought all to be zero. The statistical significance of these coefficients can be tested against a critical value to determine at what scale and translation of the wavelet the null hypothesis ought to be rejected.

[8] In this paper, we examine the stationarity of the variance of streamflow time series using the approach developed by *von Sachs and Neumann* [2000]. Using this test, we examine the stationarity characteristics of several streamflow record sets to determine if the nonstationarity (if any) of the different streamflow time series have any similarities, with regard to the time the variance changes, and on what scale the change occurs. If climate change has increased the volatility of the hydrologic cycle, we might expect to see changes in the variance structure of streamflows in separate locations occurring at the same time and on the same scale. Uncertainty about the local effects of global climate change make this hypothesis a weak one, however. It is entirely possible that climate change could lead to greater variance in some streamflows, lesser variance in some others, and no change at all in streamflow in some geographic regions. Nevertheless, an understanding of what sort of changes (if any) have occurred in streamflow may yield insight into larger questions than streamflow simulation modeling.

[9] We first outline without proofs the second-order stationarity test developed by *von Sachs and Neumann* [2000]. The test is then applied to a set of streamflow data taken from U.S. Geological Survey (USGS) gauging stations distributed across the United States. Results for the occurrence of nonstationarity in the variance of the streamflow time series are presented and discussed.

2. Data

[10] The streamflow time series used in the analysis presented in this paper were taken from USGS records of daily streamflow data. Twelve stations which had record lengths of at least 2^{14} days ($\sim$45 years) were chosen, with drainage basin areas ranging from 497 to 107,806 km^2. The records for the 2^{14} day period from 22 November 1954 to 30 September 1999 were used. A list of the stations used is given in Table 1. These gauging stations were chosen somewhat arbitrarily among the limited number of available stations with record lengths of at least 45 years, with two guiding factors: (1) a range of drainage basin sizes and (2) the



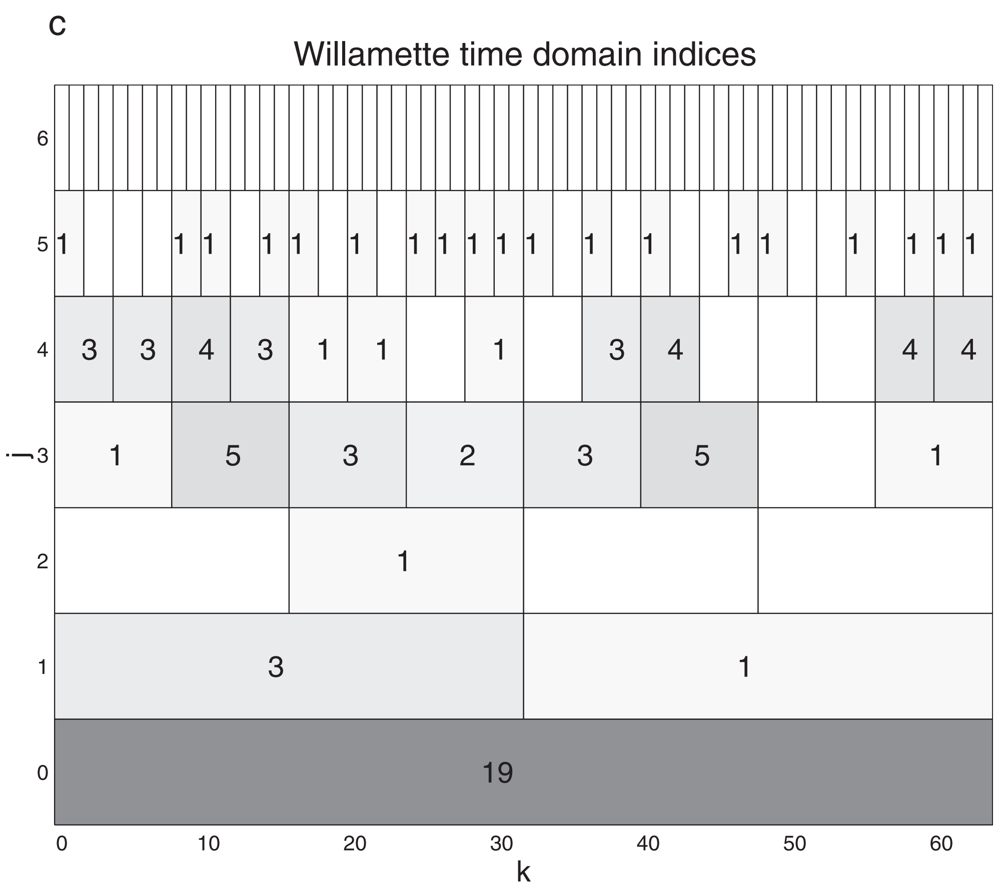


Figure 2. (continued)

desire to have different areas of the United States represented by more than one station. Total yearly flow was calculated for each river for each of the years 1955–1998, and these values are shown in Figure 1.

[11] In examining the streamflow time series a number of features were observed which would have an effect on any variance or periodogram-based analysis. If the streamflow is >100 cubic feet per second (cfs, 1 cfs = 2.8317×10^{-2} m^3/s), three significant digits are used. This means that when the flow is of the order of 10^2 cfs, the smallest reported difference between flows is 1 cfs, while when the flows are of the order of 10^4 cfs, the smallest difference between flows is 100 cfs. When the streamflow is of the order of 10 or 1 cfs, two significant digits are used, so that the smallest difference is 1 or 0.1 cfs. (Flows of the order of 1 cfs were only found in the Pecos River time series.) The fact that the data changes in discrete jumps limits the frequencies that the periodogram can resolve; perhaps more importantly, the fact that the changes in streamflow values are proportional to the amount of streamflow may introduce a heteroskedacity purely from the way the data are recorded. Of course, if the magnitude of the changes in streamflow do in fact depend on the magnitude of the streamflow (which we expect), the increase in step size as streamflow increases only leads to loss of frequency resolution. As will be discussed in Section 3, the analysis approach avoids too fine a frequency resolution for additional reasons, so that this loss of frequency resolution due to the discrete data steps should not be a significant problem.

3. Wavelet-Based Test for Variance Change

[12] As mentioned above, proofs for the statistical test outlined below are given by *Neumann and von Sachs* [1997] and *von Sachs and Neumann* [2000]. Given a time series $X_{t,\,T}(t = 1, \ldots, T)$, a time index $u \in [0, 1]$ is defined using $u = t/T$. The local periodogram (nontapered) has the form

$$I_N(u,\omega) = \frac{1}{2\pi N}\left|\sum_{s=1}^{N} X_{(uT-N/2+s),T}\exp(-i\omega s)\right|^2, \tag{2}$$

where N is a parameter defining the length of the subset of $X_{t,T}$, which is sampled for the local periodogram, the empirical estimate of $f(u,\omega)$. The time series X has spectral content over the range $[-\pi,\pi]$. In the work of *von Sachs and Neumann* [2000] the Haar wavelet in the time domain and a variation of the Littlewood-Paley wavelet in the frequency domain are used to decompose the local periodogram. These wavelets are defined as

$$\psi(u) = \begin{cases} 1 & \text{if } 0 \le u \le \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} < u \le 1 \end{cases} \tag{3}$$

and

$$\phi(\omega) = 1/\pi^{1/2} \qquad \text{for} \quad 0 \le \omega \le \pi. \tag{4}$$

The usual dyadic scaling and translation relationships for wavelets are also used:

$$\psi_{j,k}(u) = 2^{j/2}\psi\left(2^j u - k\right) \tag{5}$$

$$\phi_{j,k}(\omega) = 2^{j/2}\phi\left(2^j \omega - k\pi\right) \tag{6}$$

for $k = 0, 1, \ldots, 2^j - 1$. The dilated and translated wavelets are used to define coefficients

$$\alpha_{j,k;j',k'} = \int_0^1\int_0^\pi f(u,\omega)\psi_{j,k}(u)\phi_{j',k'}(\omega)\,du\,d\omega \tag{7}$$

$$\tilde{\alpha}_{j,k;j',k'} = \int_0^1\int_0^\pi \left[I_{\langle k2^{-j}T,(k+1/2)2^{-j}T\rangle(\omega)} + I_{\langle[(k+1/2)2^{-j}+1]T,(k+1)2^{-j}T\rangle(\omega)}\right] \cdot \psi_{j,k}(u)\phi_{j',k'}(\omega)\,du\,d\omega, \tag{8}$$

where $I_{\langle K,\,L\rangle}$ is the periodogram derived using the data from the interval $[K, L]$:

$$I_{\langle K,L\rangle}(\omega) = \frac{1}{2\pi(L-K+1)}\left|\sum_{t=K}^{L} X_t \exp(-i\omega t)\right|^2. \tag{9}$$

The indices j and k in equations (7) and (8) specify the interval $[k2^{-j}T, (k+1)\,2^{-j}T]$ over which the periodogram is computed. The index j also indicates the dilation scale of the wavelet $\psi_{j,\,k}(u)$, so that the size of this wavelet shrinks as the periodogram interval shrinks. (Equation (8) is equivalent to the first part of equation (3.1) of *von Sachs and Neumann* [2000]. The χ terms, which simply indicate the range over which the wavelet indexed by j and k is nonzero, have been removed for clarity.) With the choice of the Haar wavelet for ψ and the Littlewood-Paley wavelet for ϕ, equation (8) can be simplified to

$$\tilde{\alpha}_{j,k;j',k'} = 2^{(j+j')/2}\frac{1}{\pi^{1/2}}\int_{k'2^{-j'}\pi}^{(k'+1)2^{-j'}\pi}\left\{I_{\langle k2^{-j}T,(k+1/2)2^{-j}T\rangle(\omega)} - I_{\langle[(k+1/2)2^{-j}+1]T,(k+1)2^{-j}T\rangle(\omega)}\right\}d\omega. \tag{10}$$

The coefficient $\tilde{\alpha}_{j,k;j',k'}$ can then be seen as the difference in the periodogram for two halves of the interval $[k2^{-j}T, (k+1)\,2^{-j}T]$ for a certain frequency range $[k'2^{-j'}\pi, (k'+1)\,2^{-j'}\pi]$. If this difference is statistically significant, the periodogram and hence the variance of the time series is not stationary.

[13] The values that j and j' are allowed to take on are constrained by the length of the time series and the need for the coefficients derived in equation (10) to converge asymptotically to the theoretical value in equation (7). The diadic segment length $Nj = 2^{-(j+1)}T$ for a given value of j must satisfy $N_j \gg T^{1/2}$. This constraint applies equally to the frequency scale in j'. Given that the lengths of our time series of daily runoff were set at 2^{14} values, we therefore restricted the sum of j and j' to be less than or equal to 6 in our analysis. Hence when j is 0 (and single translation value k is also 0) and the periodogram is calculated for the entire time series $[1, T]$, the frequency scale index j' can take on any value from 0 to 6, and we have finer frequency resolution. Conversely, when $j' = 0$, j can take on any value from 0 to 6, and the resulting coefficients $\tilde{\alpha}_{j,k;0,0}$ can be applicable to temporal lengths from the entire time series ($j = 0$) to a segment of 256 records (~8 months long.)

[14] For testing statistical significance it is useful to define the variance of $\alpha_{j,k;j',\,k'}$:

$$\sigma^2_{j,k,j',k'} = 2\pi T^{-1}\int_0^1\int_0^\pi f^2(u,\omega)\psi^2_{j,k}(u)\phi^2_{j',k'}(\omega)\,du\,d\omega + o\left(T^{-1}\right) + O\left(2^{-j'}T^{-1}\right). \tag{11}$$

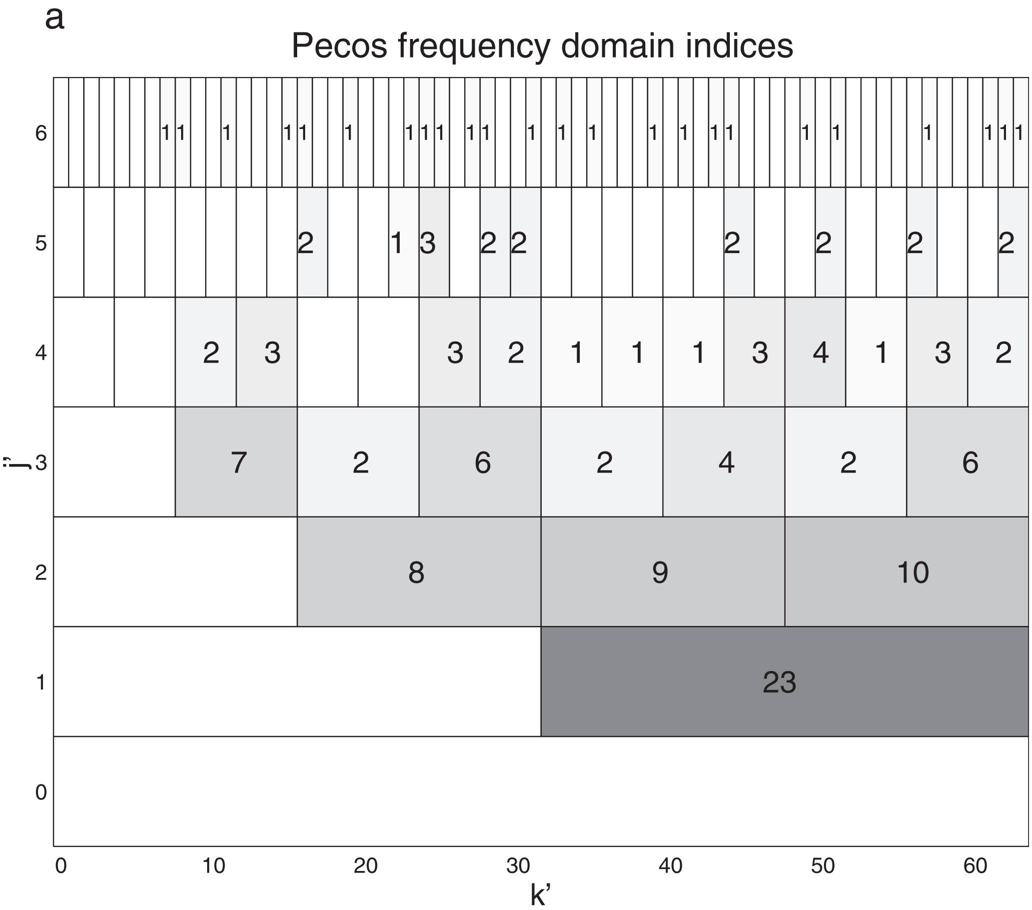


Figure 3. Plots of the number of statistically significant values of $\tilde{\alpha}_{j,k;j',k'}$ for all values of j, k (time indices) at a given set of j', k' (frequency indices) for the (a) Pecos, (b) Rapidan, and (c) Willamette Rivers. The same gray scale is used for Figures 3a–3c.

This quantity can be approximated by

$$\hat{\sigma}^2_{j,k,j',k'} = \frac{\pi}{T}\int_0^{\pi} I^2_{\langle 1,T\rangle}(\omega)\phi^2_{j',k'}(\omega)d\omega. \quad (12)$$

It can be shown [*Neumann and von Sachs*, 1997; *von Sachs and Neumann*, 2000] that the quantity

$$\frac{\tilde{\alpha}_{j,k;j',k'} - \alpha_{j,k;j',k'}}{\sigma^2_{j,k,j',k'}} \quad (13)$$

is approximately normal, so that a test of the significance of $\tilde{\alpha}_{j,k;j',k'}$ can be defined using normal critical values. Since we are evaluating the hypothesis jointly for each of $n\,\tilde{\alpha}_{j,k;j',k'}$, where n is the number of valid permutations of j, k, j', k' (769 for $j + j' \leq 6$), we adjust the chosen statistical significance level α (i.e., 5%) to a more conservative value $\alpha_T = \alpha/n$. The test then rejects the hypothesis of no variance change (which implies $H_0 : \tilde{\alpha}_{j,k;j',k'}$) if

$$\left|\tilde{\alpha}_{j,k;j',k'}\right| > \hat{\sigma}^2_{j,k,j',k'} z_{1-\alpha_T/2}, \quad (14)$$

where z is the standard normal critical value.

4. Analysis and Discussion

[15] With the constraint $j + j' \leq 6$ and $k = 0, \ldots, 2^j - 1$ and $k' = 0, \ldots, 2^{j'} - 1$, a total of 769 coefficients $\tilde{\alpha}_{j,k;j',k'}$ can be derived for each streamflow time series. Unfortunately, the fact that each coefficient has four indices associated with it means that visualization of the time-frequency relationships is not straightforward as it is with a wavelet scalogram, in which the value of the wavelet coefficient can be plotted with increasing color, indicating a larger value for the coefficient. Additionally, the actual value of $\tilde{\alpha}_{j,k;j',k'}$ is not particularly of interest, rather the fact whether it is greater than the test statistic or not is. For each set $\{j, k, j', k'\}$ where the values of j, k, j', and k' are restricted as described above, we can simply define a new variable $c_{j,k}$ which is the number of statistically significant coefficients for a given pair (j, k) for all possible (j', k'). Similarly, we can define $c'_{j',k'}$ to be the number of statistically significant coefficients for a given pair (j', k') for all possible (j, k). Since these variables depend on two indices, we can plot them in separate time and frequency plots. While the full time-frequency information is not displayed by this graphing approach, it does begin to point the way to some conclusions.

[16] These plots of $c_{j,k}$ and $c'_{j',k'}$ for the Pecos, Rapidan, and Willamette River time series are shown in Figures 2a–2c and 3a–3c. These rivers were chosen to represent the three groups of rivers selected for analysis (northwest, east, and southwest). In Figures 2a–2c the y axis indicates the value of j (the scale of the time domain wavelet). The x axis is scaled so that the diadic relationship between different scales of j is shown; that is, for $j = 0$, there is only one possible value of k (=0), so that at this level of j, the coefficients $\tilde{\alpha}_{0,0,;j',k'}$ are derived by using the entire time series, while at level 1, there are two possible values of k (=0, 1), and the coefficients $\tilde{\alpha}_{1,0,;j',k'}$ and $\tilde{\alpha}_{1,1,;j',k'}$ are each derived using half of the

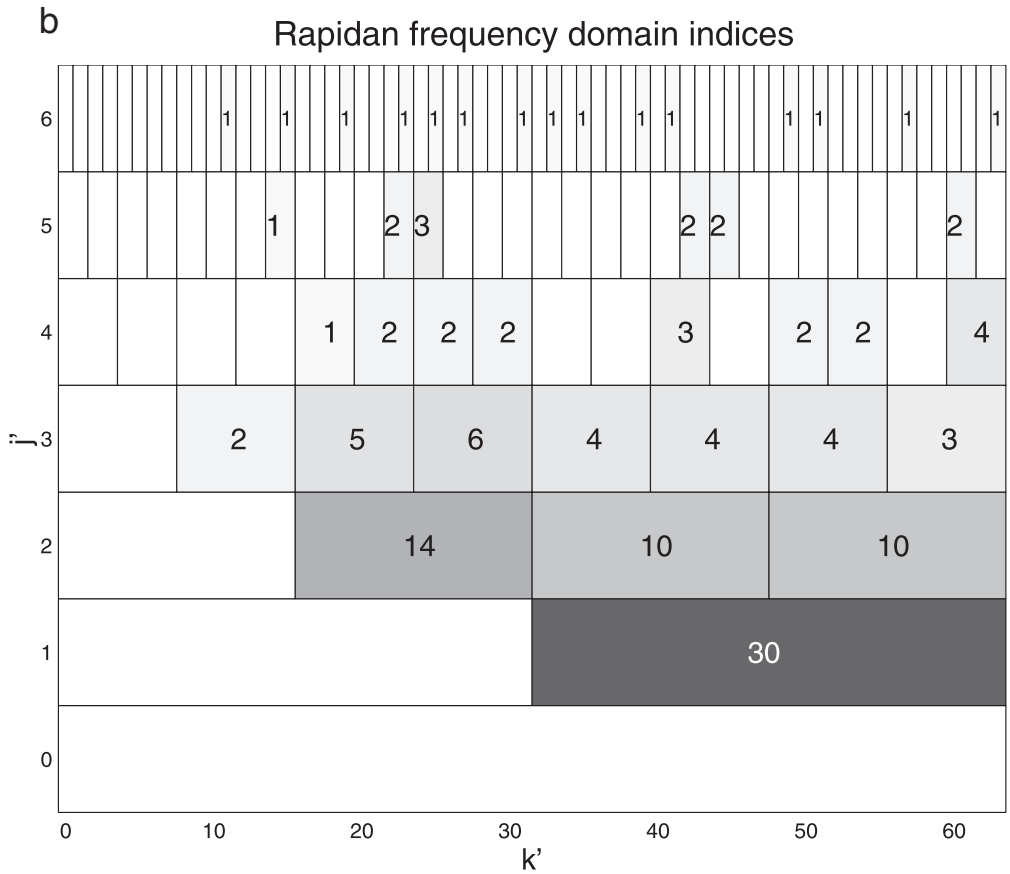


Figure 3. (continued)

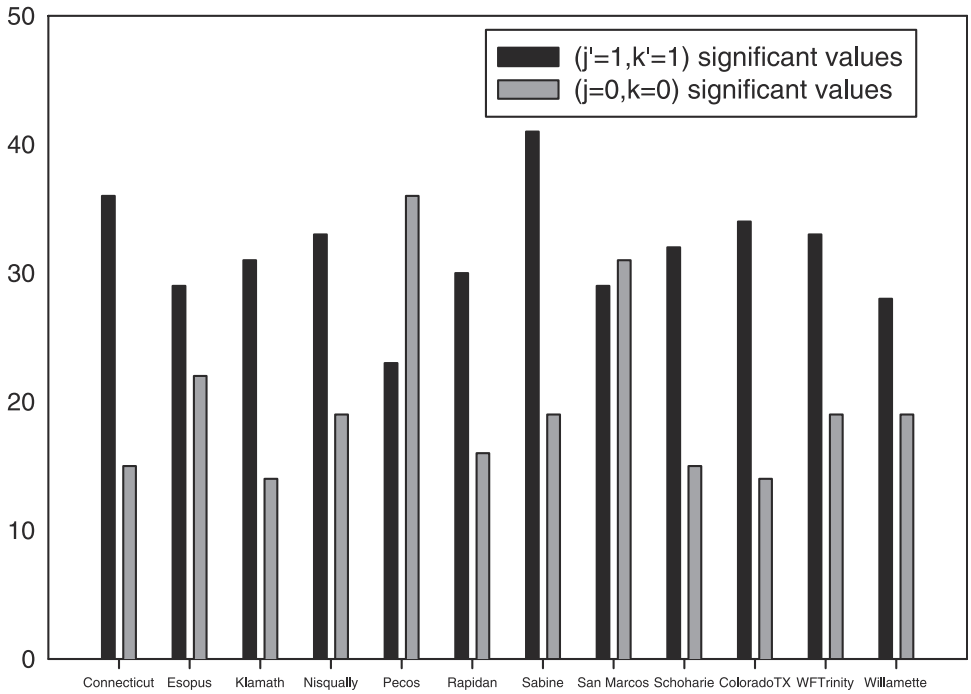


Figure 4. The number of statistically significant values of $\tilde{\alpha}_{0,0,;j',k'}$ and $\tilde{\alpha}_{j,k,1,1}$ for each river. The total number of $\tilde{\alpha}_{0,0,;j',k'}$ coefficients for each river is 127, and total number of $\tilde{\alpha}_{j,k,1,1}$ coefficients for each river is 63.

time series and so on up the levels of j. In Figures 3a–3c the same diadic plotting method is shown, except that for j' and k' we are dividing up the spectrum from 0 to π. For example, when $j' = 0$, the single resulting significance coefficient $\tilde{\alpha}_{j,k,0,0}$ is calculated from the entire periodogram from 0 to π, while when $j' = 1$, the two resulting coefficients $\tilde{\alpha}_{j,k,1,0'}$ and $\tilde{\alpha}_{j,k,1,1}$ come from the two halves of the spectrum $[0,\pi/2]$ and $[\pi/2,\pi]$. The color level indicates the number of times that $\tilde{\alpha}_{j,k;j',k'}$ was statistically significant, indicating the number of times a change in the periodogram was observed for that combination of j and k (or j' and k').

[17] It can be seen on the time plots (Figures 2a–2c) that the greatest number of ''hits'' generally occurs for the (j,k) combination (0,0), which corresponds to the periodogram for the entire time series. Only for the Rapidan River time series is this not true; for this river, $c_{0,0} = 16$, while $c_{2,1} = c_{2,3} = 17$. For the frequency plot (j',k') the combination (1,1) has the largest c' value for all three rivers shown.

[18] Although similar time and frequency plots for the other nine rivers are not shown, the same pattern of which combinations of (j,k) and (j',k') have the most statistically significant values of $\tilde{\alpha}$ holds true. For values of (j',k'), all 12 rivers have the largest number of statistically significant coefficients when $(j' = 1, k' = 1)$. Likewise, 11 of the 12 rivers have the most significant values for $(j = 0, k = 0)$; the Rapidan River is the only exception to this pattern. Figure 4 shows the number of significant coefficients for each river at $(j = 0, k = 0)$ and $(j' = 1, k' = 1)$. These values are out of 127 possible values of $c_{0,0}$ and 63 possible values of $c'_{1,1}$.

[19] Any interpretation of the hydrologic meaning of $(j = 0, k = 0)$ and $(j' = 1, k' = 1)$ having the largest number of statistically significant coefficients in the time and frequency domains should be made in light of several caveats. Different j and j' values have different numbers of possible coefficients; the effect of this is discussed below. It should be recalled that the statistical test only tests for a change in variance in the streamflow and provides no information on whether the variance increased or decreased. Finally, the following discussion only concerns the most common variance changes. As can be seen in Figures 2 and 3, variance changes occur at many different values of j,k and j',k'. All this being said, we can make some comment on the hydrologic meaning behind the results just presented. The fact that most of the rivers have the largest number of statistically significant coefficients for $(j = 0, k = 0)$ indicates that changes in variance are occurring over the entire time period under consideration (1954–1999); the variance behavior of streamflow is slowly varying. The exception to this observation is the Rapidan, which has its largest number of statistically significant coefficients at $(j = 2, k = 1)$ and $(j = 2, k = 3)$ (just slightly more significant coefficients than at $(j = 0, k = 0)$). The variance of the streamflow on the Rapidan changes during the second and fourth quarters ($\sim$1965–1976 and 1988–1999) of the time period considered here. Why the Rapidan shows variance changes more localized in time can really only be answered by consideration of other data sets, such as precipitation, that have a causal relationship to streamflow.

[20] The predominance of $(j' = 1, k' = 1)$ indicates that the most common type of variance change occurs to the higher wave number components of the streamflow. Since $(j' = 1, k' = 1)$ represents the wave number range $[\pi/2, \pi]$, which corresponds to structures of periods 2–4 days, a possible interpretation of this type of variance change is that the difference between high flow and low flow from a given storm is changing. What shows no evidence of a change in variance is the extreme lower wave number components. The combination $(j', k' = 0)$ is never statistically significant for any value of j' for any river time series. This can be interpreted as meaning that for longer-term cycles (with periods of a season or more), the variance of the streamflow is not changing, or in the limit, the average level of variation of streamflow does not change over the entire time period under consideration.

[21] This conclusion is supported by the results for $(j = 0, k = 0)$. By fixing $j = 0$ and choosing $j' = 6$, we are able to break the spectrum up into 64 bands. Nearly all the rivers only show change in the variance beginning at $k' = 11$, which corresponds to cycles in

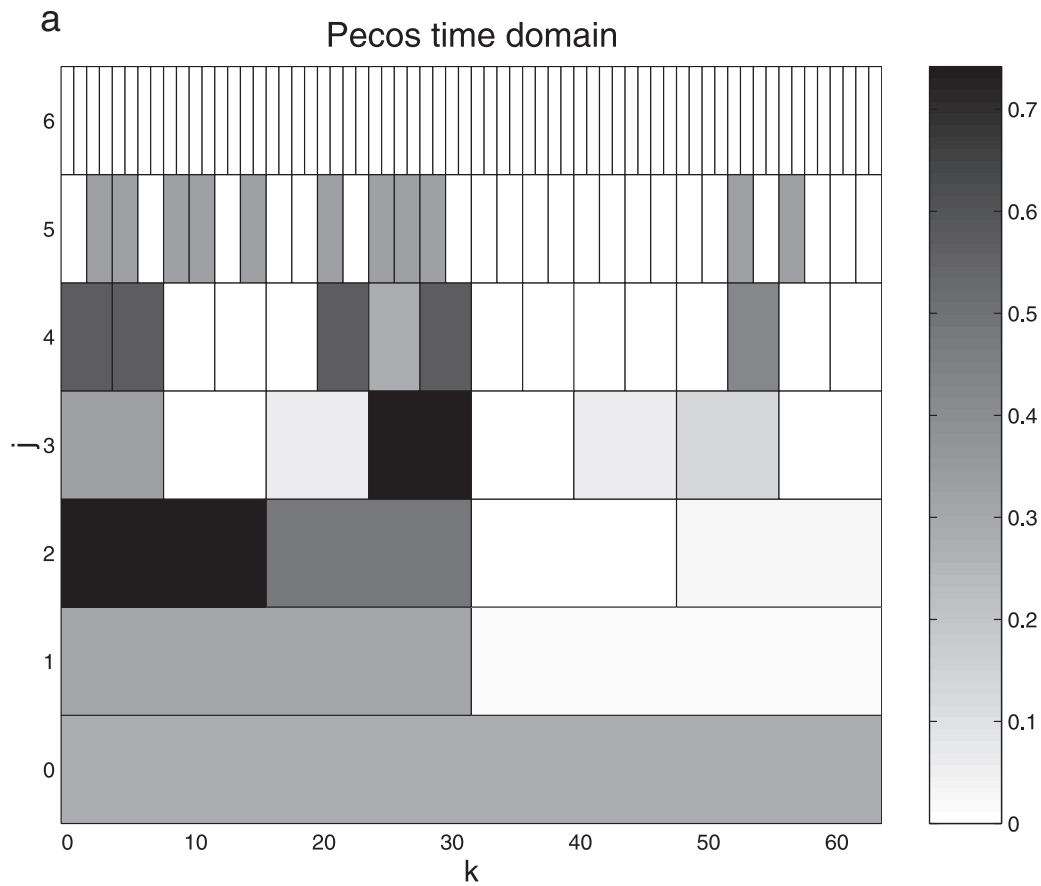


Figure 5. Graphs of the ratio $c_{j,k}/\max(c_{j,k})$ for the (a) Pecos, (b) Rapidan, and (c) Willamette Rivers. A value of 1 means that all possible coefficients $\tilde{\alpha}_{j,k;j',k'}$ are statistically significant at this value of j and k.

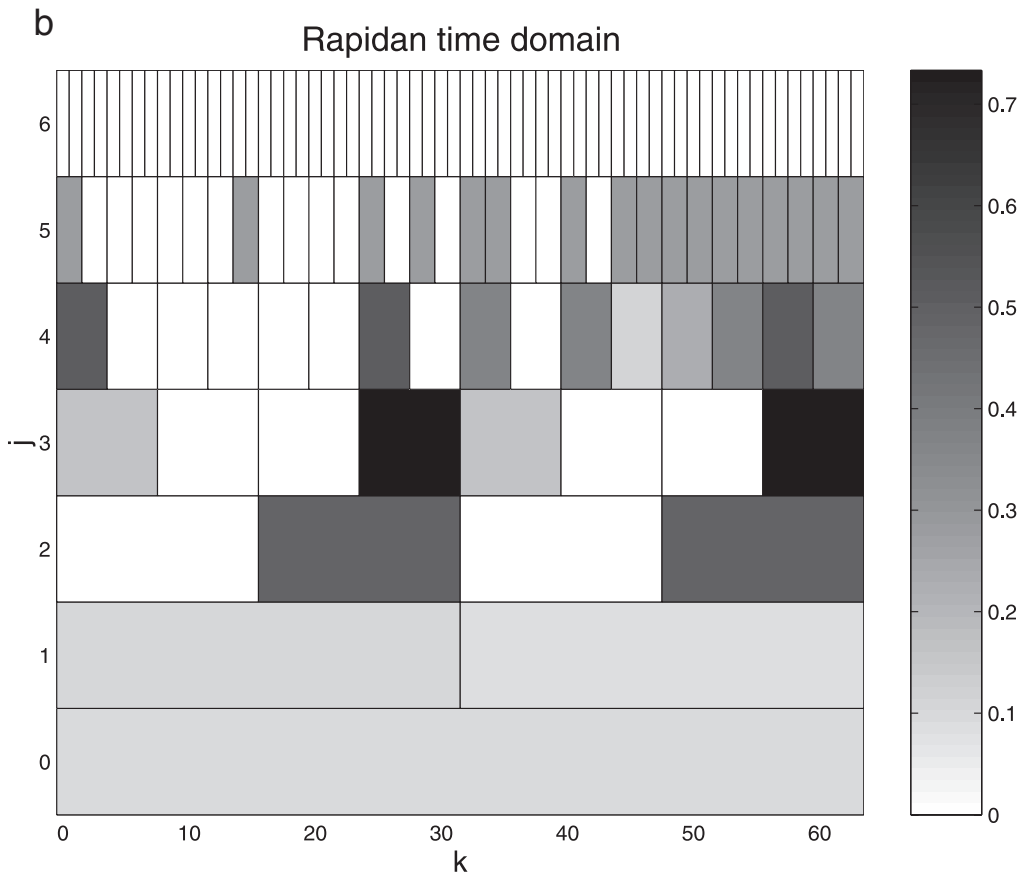
the flow of ~11 days. Only the Pecos shows any significant change at a lower wave number, for $k' = 7$, which corresponds to about a period of 18 days. In no case is any change in the variance structure seen for cycles in the streamflow on a seasonal or longer scale. The ability to resolve truly low wave number behavior is limited because we are limiting our upper value of j' to 6. Nevertheless, these results support the statement that while the behavior of the streamflow on the timescale of individual storms/runoff events may be changing, the behavior on a larger timescale (seasonally or interannually) is not.

[22] It may be pointed out that it is not unexpected that in the time plot the combination ($j = 0, k = 0$) has the most statistically significant coefficients since the value of $j = 0$ admits the most possible combinations of j' and k'. We can rescale each c value by the total number of possible statistically significant coefficients for a given combination of (j, k) or (j', k'), given $j + j' \leq 6$, so that the resulting value for every combination of (j, k) or (j', k') is between 0 and 1. Plots of the results from this rescaling for the three rivers presented previously are shown in Figures 5a–5c and 6a–6c. The problem with this sort of rescaling becomes apparent when Figures 5 and 6 are compared to Figures 2 and 3. As j or j' becomes larger and approaches the maximum value of 6, the maximum value of $c_{j,k}$ or $c'_{j',k'}$ for a given k or k' becomes smaller. Eventually, at $j = 6$, there is only one coefficient $\tilde{\alpha}_{6,k,0,0}$ for each value of k. If this coefficient is statistically significant, then the scaled value $c_{j,k}/\max(c_{j,k})$ is then 1. The same is true for $j' = 6$; the only possible coefficient for any k' is $\tilde{\alpha}_{0,0,6,k'}$. Comparisons between levels of these scaled results is made difficult by this changing of scaling factor $\max(c_{j,k})$ as the decomposition levels changes. That being said, a number of general observations can be made. Even on the rescaled frequency plots, the values for the combination ($j' = 1$, $k' = 1$) remained large compared to the values for other combinations. If the results from the levels $j' = 5$ and $j' = 6$ are not considered (since only 3 or 1 coefficients are present for each (j', k') combination at these levels), the value at ($j' = 1, k' = 1$) remains the largest for each river. In the rescaled time domain plots, for many of the rivers, large values can be seen for some combinations of ($j = 4, k$) and ($j = 3$, k).

[23] Since we have avoided the use of transformations of the time series, we ought to address the questions of trends, which would normally be removed by differencing. A linear or other trend in the data causes nonstationarity in the mean. As mentioned, previous researchers have investigated the possibility of changes in streamflow time series by testing for the existence of trends [*Changnon and Demissie*, 1996]. The existence of a linear trend in any untransformed daily river flow time series is not easy to discern because the peaks in daily discharge caused by storms act as repeated outliers that pull the slope of the regression line. Perhaps more importantly, the nature of the statistical test presented above ought to be relatively insensitive to linear trends. The test compares the difference between the periodogram calculated over two different intervals

$$I_{\langle k2^{-j}T,(k+1/2)2^{-j}T\rangle}(\omega) - I_{\langle (k+1/2)2^{-j}T,(k+1)2^{-j}T\rangle}(\omega). \qquad (15)$$

The signal over this interval can be decomposed into a linear trend and a detrended portion of the signal. The linear trend over the first



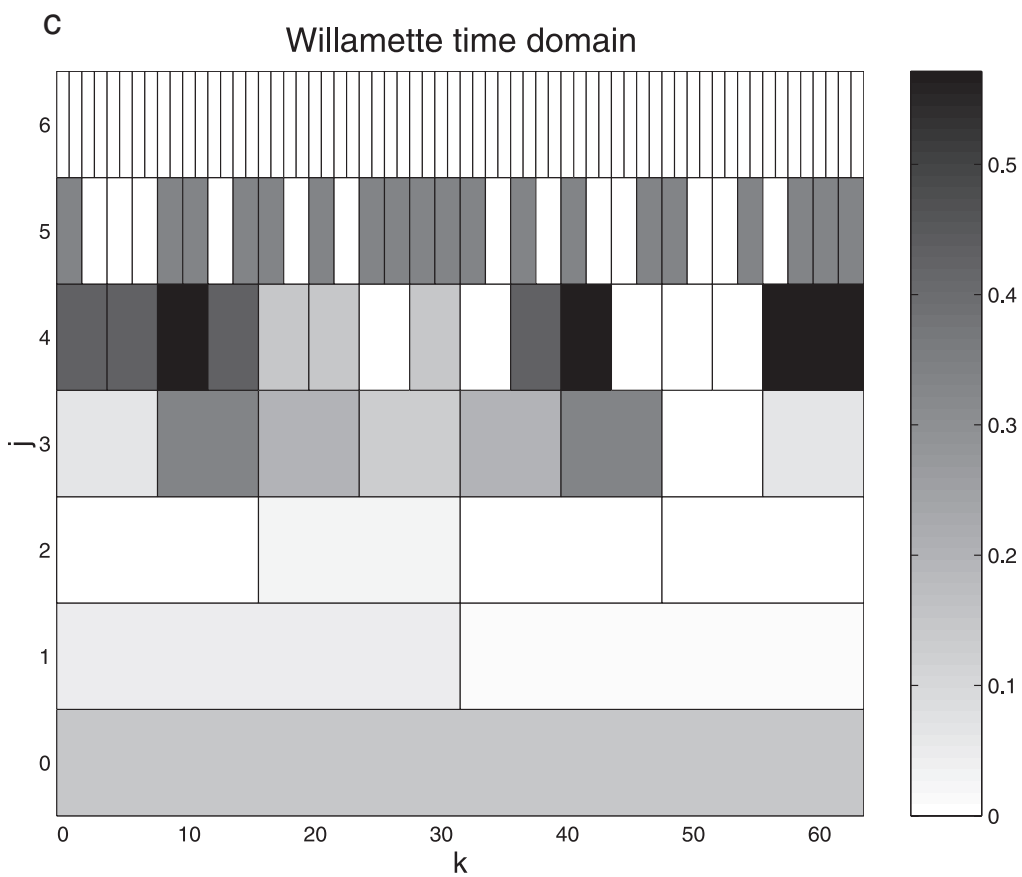


Figure 5. (continued)

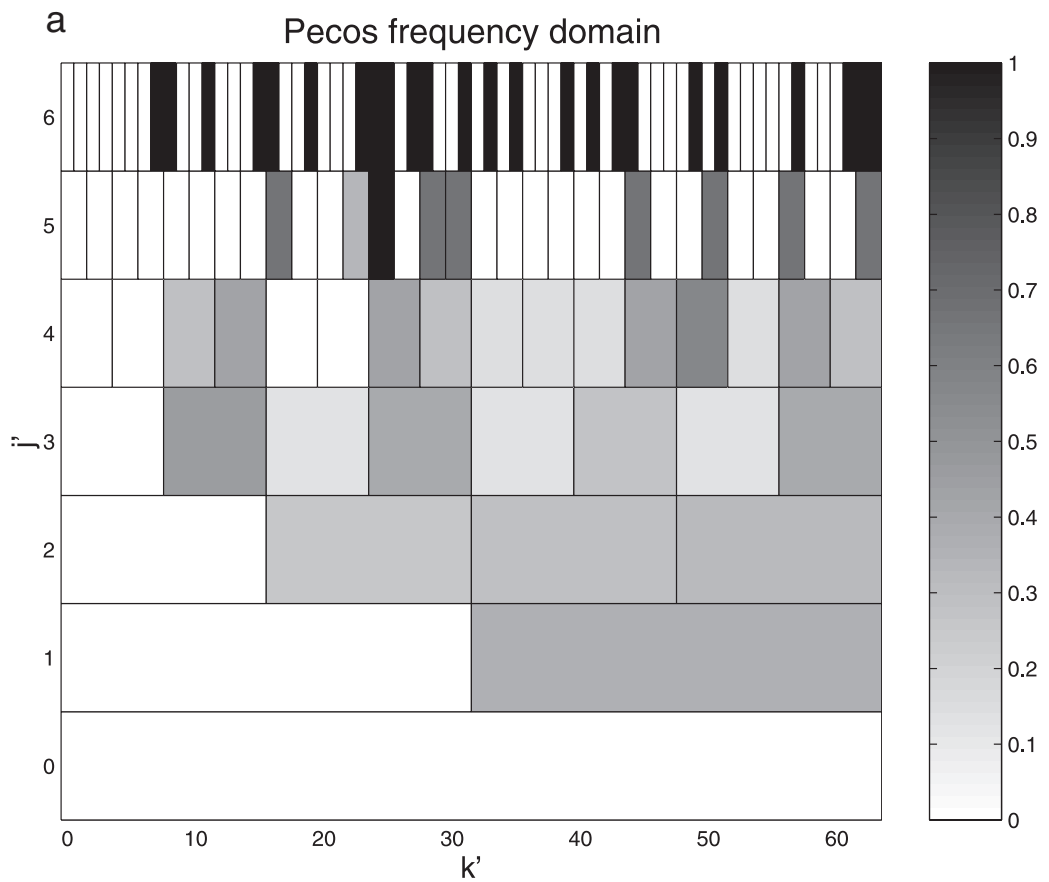


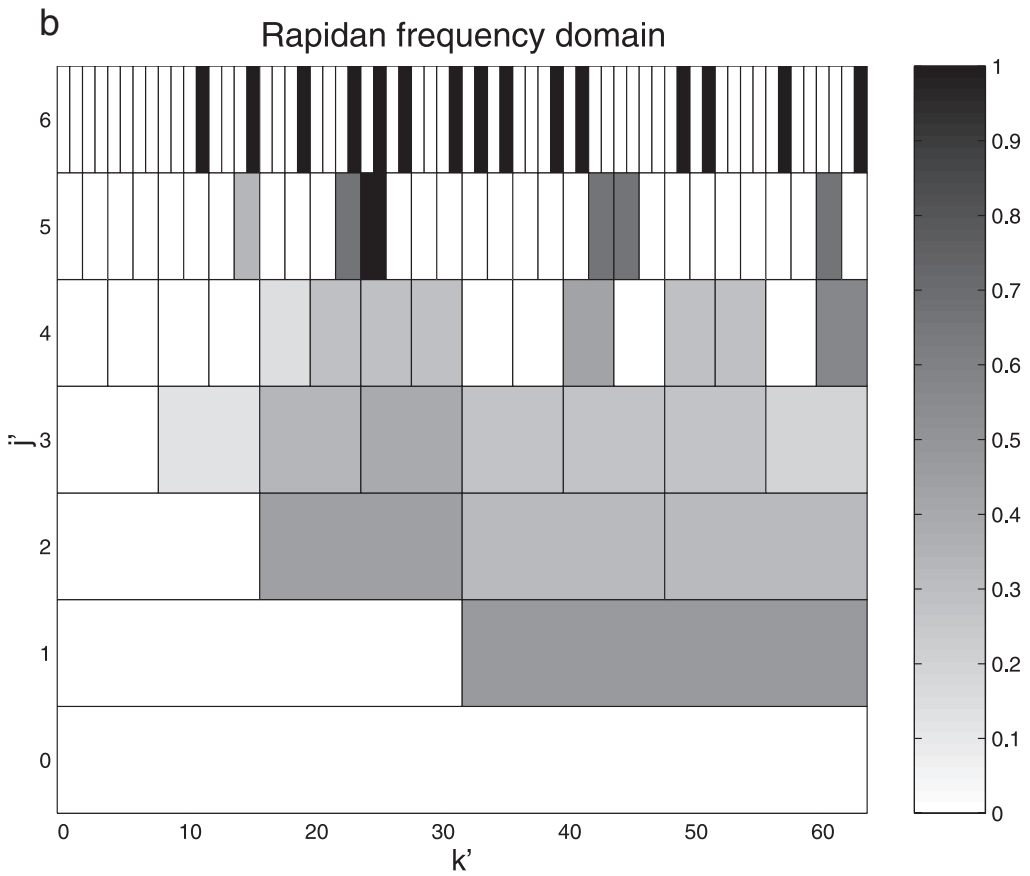
Figure 6. Graphs of the ratio $c'_{j',k'}/\max(c'_{j',k'})$ for the (a) Pecos, (b) Rapidan, and (c) Willamette Rivers. A value of 1 means that all possible coefficients $\tilde{\alpha}_{j,k;j',k'}$ are statistically significant at this value of j' and k'.

interval can be written by the set f_n, while over the second interval it is g_n, where n is simply the number of values in the interval. Since every $g_n - f_n = c$, a constant, the Fourier transforms of the linear trend in each subinterval (rescaled to 0–1) will differ only in the zeroth wave number, representing the mean of the linear trend each interval. The variance ought not to be affected by a change in mean since, by definition, the variance has the mean removed.

[24] To examine whether there was any trend in the data used, the total annual flow was computed for the years 1955–1998, and a linear regression against time was performed. A t test was then used to determine whether the regression slope was significantly different (at the 95% level) from a slope of 0. Three of the 12 streamflow time series (Connecticut, Rapidan, West Fork Trinity) were found to have statistically significant trends. The presence of these trends seems not to affect the results described above since the coefficient results for these three rivers did not differ from the other nine. The atypical behavior of $c_{j,k}$ for the Rapidan cannot be attributed simply to the presence of this annual linear trend since neither the Connecticut nor the West Fork Trinity showed similar behavior.

5. Conclusions

[25] The use of a wavelet-based test for a change in the variance of 12 streamflow time series has shown that there may have indeed been a change in the variance of daily streamflow over the past 45 years. For all 12 rivers, a large number of statistically significant results are found when a Haar wavelet scaled to the entire time duration is used in the test. Use of a frequency-limited wavelet that focused on the frequency range $[\pi/2, \pi]$ (that is, $j' = 1$, $k' = 1$) also yielded a large number of statistically significant results. Although individual river time series show significant variance changes in shorter time periods or smaller frequency bands, no consistent pattern can be seen from the results of the analysis for these 12 rivers, except for the two consistent results noted above. The fact that the most consistent results were seen using wavelets that correspond to the entire time period or large portions of the spectral band means that the test provides little evidence that allows us to localize a specific date at which the variance of streamflow changed or a specific periodic component (annual cycle, seasonal cycle, etc.) of the time series that changed. Results for the frequency case were largely confirmed when the number of statistically significant coefficients was rescaled by the total possible number of coefficients at each decomposition level. Rescaling results for the time domain were less conclusive. The particular test used in this study is one of several wavelet-based tests of the variance structure of time series and was chosen because it allowed isolation of the time-frequency ranges in the periodogram. Other wavelet-based tests of possible changes in the variance of streamflow time series may be able to pinpoint better when variance changes occur [*Whitcher et al.*, 1998] but may not provide the same frequency-based information. Resolution of when and at what frequencies the variance changes in each streamflow time series is ultimately limited in this method by the need for a long enough sample to accurately resolve the spectrum. Ideally, longer time series would allow better resolution, but because of



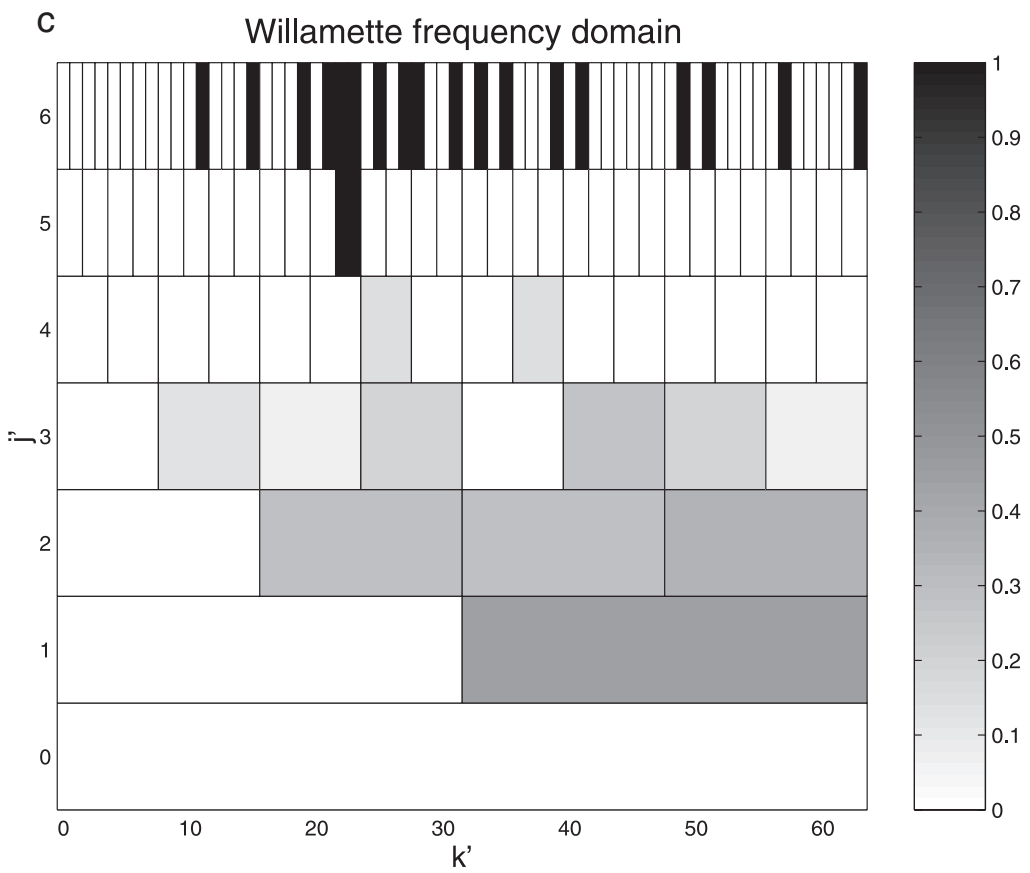


Figure 6. (continued)

diadic nature of the wavelet used in this test, the length of the time series must be a power of 2. Since the time series used in this study were 2^{14} records long, or ~45 years, the next possible time series length would be ~90 years. There are unfortunately very few daily streamflow records that long, so the resolution of when and how the variance changes presented in this paper may be close to the limit possible for streamflow time series using this method.

[26] It must be kept in mind that the changes in variance results shown here make no comment on the causes of these changes. A change in the variance of daily streamflow could arise from human-induced changes in the management of the rivers, changes in the land cover of the drainage areas, or changes in the precipitation regime in the drainage area. The fact that the rivers chosen for this study are separated widely geographically and have very different management and use regimes would seem to indicate that any management-induced changes in the rivers ought to be decorrelated. The question of how the results from one river relate to the results from another could be potentially pursued by incorporating information on land use changes and climatic change over time into a larger analysis. A coupling of the statistical analysis of streamflow time series and time series of forcing variables such as land cover change and/or precipitation and other climatic variables would mimic more closely the analysis done in the forward modeling problem and would provide more complete answers.

References

Beran, J., Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models, *J. R. Stat. Soc., Ser. B*, *57*, 659–672, 1995.

Box, G. E. P., and D. R. Cox, An analysis of transformations, *J. R. Stat. Soc., Ser. B*, *26*, 211–252, 1964.

Bras, R. L., and I. Rodriguez-Iturbe, *Random Functions and Hydrology*, Dover, Mineola, N. Y., 1993.

Brutsaert, W., and M. B. Parlange, Hydrologic cycle explains the evaporation paradox, *Nature*, *396*, 30, 1998.

Changnon, S. A., and M. Demissie, Detection of changes in streamflow and floods resulting from climate fluctuations and land-use changes, *Clim. Change*, *32*, 411–421, 1996.

Coifman, R. R., and D. Donoho, Translation-invariant denoising, in *Wavelets and Statistics*, edited by A. Antoniadis and G. Oppenheim, *Lect. Notes Stat.*, *103*, 125–150, 1995.

Hamlet, A. F., and D. P. Lettenmaier, Effects of climate change on hydrology and water resources in the Columbia River basin, *J. Am. Water Resour. Assoc.*, *35*, 1597–1623, 1999.

Inclán, C., and G. C. Tiao, Use of cumulative sumes of squares for retrospective detection of changes of variance, *J. Am. Stat. Assoc.*, *89*, 913–923, 1994.

Jiang, J., D. Zhang, and K. Fraedrich, Historic climate variability of wetness in east China (960-1992): A wavelet analysis, *Int. J. Climatol.*, *17*, 969–981, 1997.

Johnson, R. A., and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall, Old Tappan, N. J., 1998.

Karl, T. R., R. W. Knight, D. R. Easterling, and R. G. Quayle, Indices of climate change for the United States, *Bull. Am. Meteorol. Soc.*, *77*, 279–292, 1996.

Katul, G. G., and M. B. Parlange, Analysis of land surface heat fluxes using the orthonormal wavelet approach, *Water Resour. Res.*, *31*, 2743–2749, 1995.

Kiely, G., J. D. Albertson, and M. B. Parlange, Recent trends in diurnal variation of precipitation at Valentia on the west coast of Ireland, *J. Hydrol.*, *207*, 270–279, 1998.

Kumar, P., and E. Foufoula-Georgiou, A multicomponent decomposition of spatial rainfall fields, 1, Segregation of large- and small-scale features using wavelet transforms, *Water Resour. Res.*, *29*, 2515–2532, 1993.

Li, T.-H., Time correlation analysis of nonstationary time series, *J. Time Ser. Anal.*, *19*, 47–67, 1998.

Liu, Q., S. Islam, I. Rodriguez-Iturbe, and Y. Le, Phase-space analysis of daily streamflow: Characterization and prediction, *Adv. Water Res.*, *21*, 463–475, 1998.

Mallat, S., G. Papanicolaou, and Z. Zhang, Adaptive covariance estimation of locally stationary processes, *Ann. Stat.*, *26*, 1–47, 1998.

Matheussen, B., R. L. Kirschbaum, I. A. Goodman, G. M. O'Donnell, and D. P. Lettenmaier, Effects of land cover change on streamflow in the interior Columbia River Basin (USA and Canada), *Hydrol. Process.*, *14*(5), 867–885, 2000.

McCulloch, R. E., and R. S. Tsay, Bayesian inference and prediction for mean and variance shifts in autoregressive time series, *J. Am. Stat. Assoc.*, *88*, 968–978, 1993.

Neumann, M. H., and R. von Sachs, Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra, *Ann. Stat.*, *25*, 38–76, 1997.

Percival, D. B., On estimation of the wavelet variance, *Biometrika*, *82*, 619–631, 1995.

Percival, D. B., and H. O. Mofjeld, Analysis of subtidal coastal sea level fluctuations using wavelets, *J. Am. Stat. Assoc.*, *92*, 868–880, 1997.

Percival, D. B., and A. T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge Univ. Press, New York, 2000.

Picard, D., Testing and estimating change-points in time series, *Adv. Appl. Prob.*, *17*, 841–867, 1985.

Smith, L. C., D. L. Turcotte, and B. L. Isacks, Stream flow characterization and feature detection using a discrete wavelet transform, *Hydrol. Proc.*, *12*, 233–249, 1998.

von Sachs, R., and M. H. Neumann, A wavelet-based test for stationarity, *J. Time Ser. Anal.*, *21*, 597–613, 2000.

Wang, B., and Y. Wang, Temporal structure of the Southern Oscillation as revealed by waveform and wavelet analysis, *J. Clim.*, *9*, 1586–1598, 1996.

Whitcher, B., S. D. Byers, P. Guttorp, and D. B. Percival, Testing for homogeneity of variance in time series: Long memory, wavelets and the Nile River, *Tech. Rep. 009*, Natl. Res. Cent. for Stat. and the Environ., Univ. of Wash., Seattle, 1998.

Whitcher, B., P. Guttorp, and D. B. Percival, Wavelet analysis of covariance with application to atmospheric time series, *J. Geophys. Res.*, *105*(D11), 14,941–14,962, 2000.

Yiou, P., D. Sornette, and M. Ghil, Data-adaptive wavelets and multiscale singular-spectrum analysis, *Physica D*, *142*, 254–290, 2000.

A. T. Cahill, Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843, USA.